

AD-A101 851

CALIFORNIA UNIV BERKELEY OPERATIONS RESEARCH CENTER

F/G 12/1

#### **MULTI-SERVER QUEUES. (U)**

JUN 81 S M ROSS

ORC-81-17

100

AFOSR-81-0122

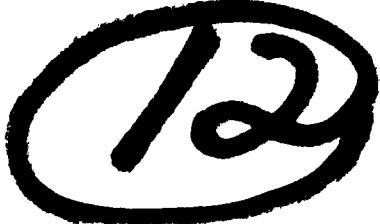
NL

UNCLASSIFIED

END  
DATE  
FILED  
H B  
OTIC

ORC 81-17  
JUNE 1981

LEVEL II



MULTI-SERVER QUEUES

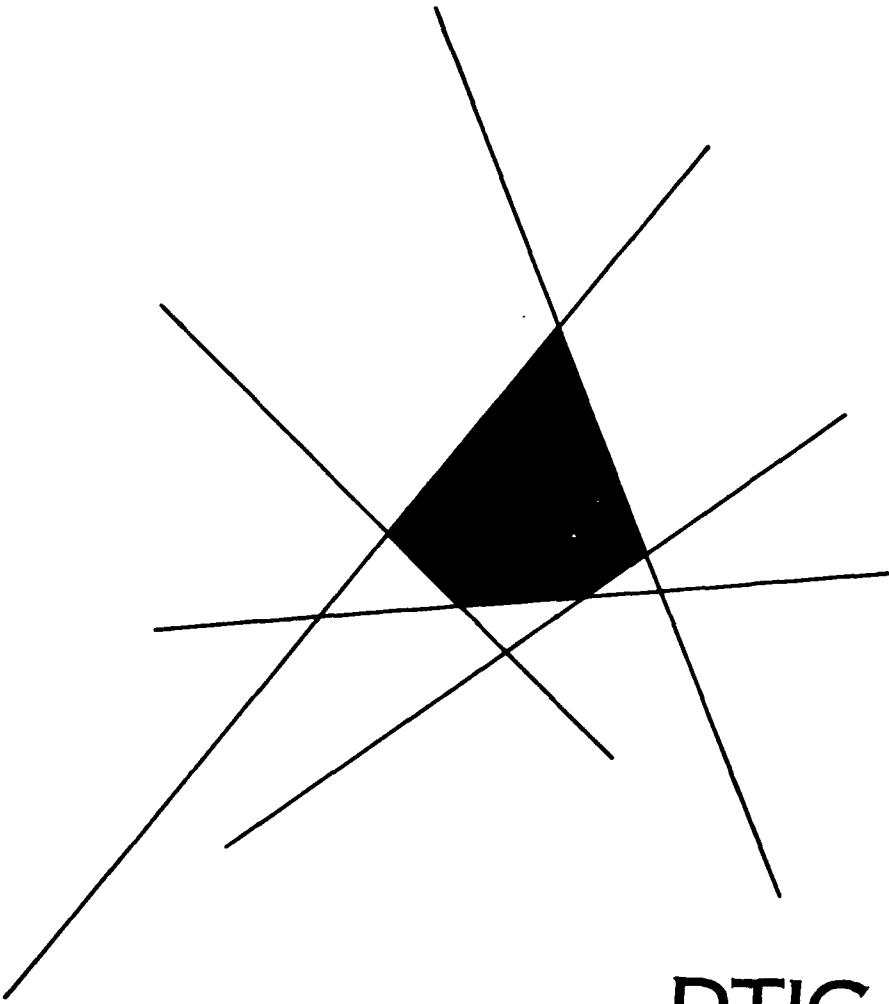
by

SHELDON M. ROSS

ADA101851

516 FILE COPY

OPERATIONS  
RESEARCH  
CENTER



DTIC  
ELECTED  
S JUL 23 1981 D  
D

DISTRIBUTION STATEMENT A  
Approved for public release;  
Distribution Unlimited

UNIVERSITY OF CALIFORNIA · BERKELEY

81 7 22 054

<b>Accession For</b>	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
<b>Justification</b>	
By _____	
Distribution/ _____	
Availability Codes _____	
Dist	Avail and/or Special
A	

MULTI-SERVER QUEUES

by

Sheldon M. Ross  
 Department of Industrial Engineering  
 and Operations Research  
 University of California, Berkeley

DTIC  
 ELECTE  
 S JUL 23 1981 D  
 D

JUNE 1981

ORC 81-17

This research was supported by the Air Force Office of Scientific Research (AFSC), USAF, under Grant AFOSR-81-0122 with the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government.

DISTRIBUTION STATEMENT A
Approved for public release; Distribution Unlimited

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER <i>(14) ORC-81-17</i>	2. GOVT ACCESSION NO. <i>AD-A202 852</i>	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) <i>(4) MULTI-SERVER QUEUES</i>	5. TYPE OF REPORT & PERIOD COVERED <i>(9) Research Report</i>	
6. PERFORMING ORG. REPORT NUMBER		
7. AUTHOR(s) <i>(10) Sheldon M. Ross</i>	8. CONTRACT OR GRANT NUMBER(s) <i>(13) AFOSR-81-0122✓</i>	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Operations Research Center University of California Berkeley, California 94720	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS <i>(11) 2304/A5 (11)</i>	
11. CONTROLLING OFFICE NAME AND ADDRESS United States Air Force Air Force Office of Scientific Research Bolling Air Force Base, D.C. 20332	12. REPORT DATE <i>(11) June 1981</i>	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	13. NUMBER OF PAGES <i>(11) 16</i>	
16. DISTRIBUTION STATEMENT (of this Report)  Approved for public release; distribution unlimited.	15. SECURITY CLASS. (of this report) Unclassified	
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Multiserver Erlang Loss Processor Sharing Finite Capacity M/G/k G/M/k	Reversed Process	
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)  (SEE ABSTRACT)		

## MULTI-SERVER QUEUES

Sheldon M. Ross

Department of Industrial Engineering and  
Operations Research

### ABSTRACT

We will survey a variety of multiserver models in which the arrival stream is a Poisson process. In particular, we will consider the Erlang loss model in which arrivals finding all servers busy are lost. In this system, we assume a general service distribution. We will also consider finite and infinite capacity versions of this model. Another model of this type is the shared processor system in which service is shared by all customers.

Another model to be considered is the G/M/k in which arrivals are in accordance with a renewal process and the service distribution is exponential. We will analyze this model by means of the embedded Markov chain approach.

### 0. INTRODUCTION

We will consider some multiserver queueing models. In Section 1, we deal with the Erlang loss model which supposes Poisson arrivals and a general service distribution  $G$ . By use of a "reversed process" argument (see [2]) we will indicate a proof of the well-known result that the distribution of number of busy servers depends on  $G$  only through its mean. In Section 2 we then analyze a shared-processor model in which the servers are able to combine forces. Again making use of the reverse process, we obtain the limiting distribution for this model. In Section 3 we review the embedded Markov chain approach for the G/M/k model; and in the final section we present the model M/G/k.

### 1. THE ERLANG LOSS SYSTEM

One of the most basic types of queueing system are the loss systems in which an arrival that finds all servers busy is presumed lost to the system. The simplest such system is the M/M/k loss system in which customers arrive according to a Poisson process having rate  $\lambda$ , enter the system if at least one of the k servers is free, and then spend an exponential amount of time with rate  $\mu$  being served. The balance equations for the stationary probabilities are

$$\begin{array}{ll} \text{State} & \text{Rate leave = rate enter} \\ 0 & \lambda P_0 = \mu P_1 \\ i, 0 < i < k & (\lambda + i\mu)P_i = (i+1)\mu P_{i+1} + \lambda P_{i-1} \\ k & k\mu P_k = \lambda P_{k-1} . \end{array}$$

Using the equation  $\sum_0^k P_i = 1$ , the above equations can be solved to give

$$P_i = \frac{(\lambda/\mu)^i / i!}{\sum_{j=0}^k (\lambda/\mu)^j / j!}, \quad i = 0, 1, \dots, k.$$

Since  $E[S] = 1/\mu$ , where  $E[S]$  is the mean service time, the above can be written as

$$P_i = \frac{(\lambda E[S])^i / i!}{\sum_{j=0}^k (\lambda E[S])^j / j!}, \quad i = 0, 1, \dots, k.$$

The above was originally obtained by Erlang who then conjectured that it was valid for an arbitrary service distribution. We shall present a proof of this result, known as the Erlang loss formula when the service distribution  $G$  is continuous and has density  $g$ .

Theorem 1. The limiting distribution of the number of customers in the Erlang loss system is given by

$$P\{n \text{ in system}\} = \frac{(\lambda E[S])^n / n!}{\sum_{i=0}^k (\lambda E[S])^i / i!}, \quad n = 0, 1, \dots, k$$

and given that there are  $n$  in the system, the ages (or the residual times) of these  $n$  are independent and identically distributed according to the equilibrium distribution of  $G$ .

Proof. We can analyze the above system as a Markov process by letting the state at any time be the ordered ages of the customers in service at that time. That is, the state will be  $\underline{x} = (x_1, x_2, \dots, x_n)$ ,  $x_1 \leq x_2 \leq \dots \leq x_n$ , if there are  $n$  customers in service, the most recent one having arrived  $x_1$  time units ago, the next most recent arrival being  $x_2$  time units ago, and so on. The process of successive states will be a Markov process in the sense that the conditional distribution of any future state given the present and all the past states will depend only on the present state. In addition, let us denote by  $\lambda(t) = g(t)/\bar{G}(t)$  the hazard rate function of the service distribution.

We will attempt to use the reverse process to obtain the limiting probability density  $p(x_1, x_2, \dots, x_n)$ ,  $1 \leq n \leq k$ ,  $x_1 \leq x_2 \leq \dots \leq x_n$ , and  $P(\phi)$  the limiting probability that the system is empty. Now since the age of a customer in service increases linearly from 0 upon its arrival to its service time upon its departure, it is clear that if we look backwards, we will be following the excess or additional service time of a customer. As there will never be more than  $k$  in the system, we make the following conjecture.

Conjecture. In steady state, the reverse process is also a  $k$  server loss system with service distribution  $G$  in which arrivals occur according to a Poisson process with rate  $\lambda$ . The state at any time represents the ordered residual service times of customers presently in service. In addition, the limiting probability density is

$$p(x_1, \dots, x_n) = \frac{\lambda^n \prod_{i=1}^n \bar{G}(x_i)}{\sum_{i=0}^k (\lambda E[S])^i / i!}, \quad x_1 \leq x_2 \leq \dots \leq x_n$$

and

$$P(\phi) = \left[ \sum_{i=0}^k (\lambda E[S])^i / i! \right]^{-1}.$$

To verify the conjecture, for any state  $\underline{x} = (x_1, \dots, x_i, \dots, x_n)$ , let  $e_i(\underline{x}) = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ . Now in the original process when the state is  $\underline{x}$ , it will instantaneously go to  $e_i(\underline{x})$  with a probability density equal to  $\lambda(x_i)$  since the

person whose time in service is  $x_i$  would have to instantaneously complete its service. Similarly in the reversed process if the state is  $e_i(\underline{x})$ , then it will instantaneously go to  $\underline{x}$  if a customer having service time  $x_i$  instantaneously arrives. So we see that

in forward:  $\underline{x} \rightarrow e_i(\underline{x})$  with probability intensity  $\lambda(x_i)$   
 in reverse:  $e_i(\underline{x}) \rightarrow \underline{x}$  with (joint) probability intensity  $\lambda g(x_i)$ .

Hence if  $p(\underline{x})$  represents the limiting density, then we would need that

$$p(\underline{x})\lambda(x_i) = p(e_i(\underline{x}))\lambda g(x_i)$$

or, since  $\lambda(x_i) = g(x_i)/\bar{G}(x_i)$ ,

$$p(\underline{x}) = p(e_i(\underline{x}))\lambda \bar{G}(x_i)$$

which is easily seen to be satisfied by the conjectured  $p(\underline{x})$ .

To complete our proof of the conjecture, we must consider transitions of the forward process from  $\underline{x}$  to  $(0, \underline{x}) = (0, x_1, x_2, \dots, x_n)$  when  $n < k$ . Now

in forward:  $\underline{x} \rightarrow (0, \underline{x})$  with instantaneous intensity  $\lambda$   
 in reverse:  $(0, \underline{x}) \rightarrow \underline{x}$  with probability 1.

Hence we must verify that

$$p(\underline{x})\lambda = p(0, \underline{x})$$

which easily follows since  $\bar{G}(0) = 1$ .

Hence we see that the conjecture is true and so, upon integration, we obtain

$P\{n \text{ in the system}\}$

$$= P(\phi)\lambda^n \int_{x_1 \leq x_2 \leq \dots \leq x_n} \prod_{i=1}^n \bar{G}(x_i) dx_1 dx_2 \dots dx_n$$

$$= P(\phi) \frac{\lambda^n}{n!} \int_{x_1, x_2, \dots, x_n} \prod_{i=1}^n \bar{G}(x_i) dx_1 dx_2 \dots dx_n$$

$$= P(\phi)(\lambda E[S])^n / n!, \quad n = 1, 2, \dots, k$$

where  $E[S] = \int \bar{G}(x)dx$  is the mean service time. Also, we see that the conditional distribution of the ordered ages given that there are  $n$  in the system is

$$p(\underline{x} | n \text{ in the system}) = p(\underline{x}) / P\{n \text{ in the system}\}$$

$$= n! \prod_{i=1}^n (\bar{G}(x_i)/E[S]).$$

As  $\bar{G}(x)/E[S]$  is just the density of  $G_e$ , the equilibrium distribution of  $G$ , this completes the proof. ||

In addition, by looking at the reversed process, we also have the following corollary.

Corollary 1. In the Erlang loss model, the departure process (including both customers completing service and those that are lost) is a Poisson process at rate  $\lambda$ .

Proof. The above follows since in the reversed process arrivals of all customers (including those that are lost) constitutes a Poisson process.

## 2. THE SHARED PROCESSOR SYSTEM

Suppose that customers arrive in accordance with a Poisson process having rate  $\lambda$ . Each customer requires a random amount of work, distributed according to  $G$ . The server can process work at a rate of one unit of work per unit time, and divides his time equally among all of the customers presently in the system. That is, whenever there are  $n$  customers in the system, each will receive service work at a rate of  $1/n$  per unit time.

Let  $\lambda(t)$  denote the failure rate function of the service distribution, and suppose that  $\lambda E[S] < 1$  where  $E[S]$  is the mean of  $G$ .

To analyze the above, let the state at any time be the ordered vector of the amounts of work already performed on customers still in the system. That is, the state is  $\underline{x} = (x_1, x_2, \dots, x_n)$ ,  $x_1 \leq x_2 \leq \dots \leq x_n$  if there are  $n$  customers in the system and  $x_1, \dots, x_n$  is the amount of work performed on these  $n$  customers. Let  $p(\underline{x})$  and  $P(\phi)$  denote the limiting probability density and the limiting probability that the system is empty. We make the following conjecture regarding the reverse process.

Conjecture. The reverse process is a system of the same

type, with customers arriving at a Poisson rate  $\lambda$ , having workloads distributed according to  $G$  and with the state representing the ordered residual workloads of customers presently in the system.

To verify the above conjecture and at the same time obtain the limiting distribution let  $e_i(\underline{x}) = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$  when  $\underline{x} = (x_1, \dots, x_n)$ ,  $x_1 \leq x_2 \leq \dots \leq x_n$ . Note that

in forward:  $\underline{x} \rightarrow e_i(\underline{x})$  with probability intensity  $\frac{\lambda(x_i)}{n}$   
 in reverse:  $e_i(\underline{x}) \rightarrow \underline{x}$  with (joint) probability intensity  $\lambda G'(x_i)$ .

The above follows as in the previous section with the exception that if there are  $n$  in the system then a customer who already had the amount of work  $x_i$  performed on it will instantaneously complete service with probability  $\lambda(x_i)/n$ .

Hence, if  $p(\underline{x})$  is the limiting density then we need that

$$p(\underline{x}) \frac{\lambda(x_i)}{n} = p(e_i(\underline{x})) \lambda G'(x_i)$$

or, equivalently,

$$\begin{aligned} p(\underline{x}) &= n \bar{G}(x_i) p(e_i(\underline{x})) \lambda \\ &= n \bar{G}(x_i) (n - 1) \bar{G}(x_j) p(e_j(\underline{x})) \lambda^2, \quad i \neq j \\ &\vdots \\ &= n! \lambda^n P(\phi) \prod_{i=1}^n \bar{G}(x_i). \end{aligned} \tag{1}$$

Integrating over all vectors  $\underline{x}$  yields

$$P\{n \text{ in system}\} = (\lambda E[S])^n P(\phi).$$

Using

$$P(\phi) + \sum_{n=1}^{\infty} P\{n \text{ in the system}\} = 1$$

gives

$$P\{n \text{ in the system}\} = (\lambda E[S])^n (1 - \lambda E[S]), \quad n \geq 0.$$

Also, the conditional distribution of the ordered amounts of

work already performed, given  $n$  in the system is, from (1)

$$p(\underline{x} \mid n) = p(\underline{x}) / P\{n \text{ in system}\}$$

$$= n! \prod_{i=1}^n (\bar{G}(x_i) / E[S]) .$$

That is, given  $n$  customers in the system the unordered amount of work already performed are distributed independently according to  $G_e$ , the equilibrium distribution of  $G$ .

All of the above is based on the assumption that the conjecture is valid. To complete the proof of its validity, we must verify that

$$p(\underline{x})\lambda = p(0, \underline{x}) \frac{1}{n+1} .$$

The above being the relevant equation since the reverse process when in state  $(\epsilon, \underline{x})$  will go to state  $\underline{x}$  in time  $(n+1)\epsilon$ . As the above is easily verified, we have thus shown

Theorem 2. For the Processor Sharing Model, the number of customers in the system has the distribution

$$P\{n \text{ in system}\} = (\lambda E[S])^n (1 - \lambda E[S]) , n \geq 0 .$$

Given  $n$  in the system, the completed (or residual) workloads are independent and have distribution  $G_e$ . The departure process is a Poisson process with rate  $\lambda$ .

If we let  $L$  denote the average number in the system, and  $W$ , the average time a customer spends in the system then

$$\begin{aligned} L &= \sum_{n=0}^{\infty} n(\lambda E[S])^n (1 - \lambda E[S]) \\ &= \frac{\lambda E[S]}{1 - \lambda E[S]} . \end{aligned}$$

We can obtain  $W$  from the well-known formula  $L = \lambda W$  and so

$$W = L/\lambda = \frac{E[S]}{1 - \lambda E[S]} .$$

Another interesting computation in this model is that of the conditional mean time an arrival spends in the system given its workload is  $y$ . To compute this quantity, fix  $y$  and say that a customer is "special" if its workload is between  $y$  and

$y + \epsilon$ . By  $L = \lambda W$ , we thus have that

$$\begin{aligned} & \text{Average Number of Special Customers in the System} \\ &= \text{Average Arrival Rate of Special Customer} \times \text{Average} \\ & \quad \text{Time a Special Customer Spends in the System.} \end{aligned}$$

To determine the average number of special customers in the system, let us first determine the density of the total workload of an arbitrary customer presently in the system. Suppose such a customer has already received the amount of work  $x$ . Then the conditional density of its workload is

$$f(w \mid \text{has received } x) = g(w)/\bar{G}(x), \quad x \leq w.$$

But, from Theorem 2, the amount of work an arbitrary customer in the system has already received has the distribution  $G_e$ . Hence the density of the total workload of someone present in the system is

$$\begin{aligned} f(w) &= \int_0^w \frac{g(x)}{\bar{G}(x)} dG_e(x) \\ &= \int_0^w \frac{g(x)}{E[S]} dx, \quad \text{since } dG_e(x) = \bar{G}(x)/E[S] \\ &= wg(w)/E[S]. \end{aligned}$$

Hence the average number of special customers in the system is

$$\begin{aligned} & E[\text{number in system having workload between } y \text{ and } y + \epsilon] \\ &= Lf(y)\epsilon + o(\epsilon) \\ &= Lyg(y)\epsilon/E[S] + o(\epsilon). \end{aligned}$$

In addition, the average arrival rate of customers whose workload is between  $y$  and  $y + \epsilon$  is

$$\text{Average arrival rate} = \lambda g(y)\epsilon + o(\epsilon).$$

Hence we see that

$$\begin{aligned} & E[\text{time in system} \mid \text{workload in } (y, y + \epsilon)] \\ &= \frac{Lyg(y)\epsilon}{E[S]\lambda g(y)\epsilon} + \frac{o(\epsilon)}{\epsilon}. \end{aligned}$$

Letting  $\varepsilon \rightarrow 0$ , we obtain

$$\begin{aligned} E[\text{time in system} \mid \text{workload is } y] &= \frac{y}{\lambda E[S]} L \\ &= \frac{y}{1 - \lambda E[S]} . \end{aligned}$$

Thus the average time in the system of a customer needing  $y$  units of work also depends on the service distribution only through its mean.

### 3. THE G/M/k QUEUE

In this model we suppose that there are  $k$  servers, each of whom serves at an exponential rate  $\mu$ . We allow the time between successive arrivals to have an arbitrary distribution  $G$ . In order to ensure that a steady-state (or limiting) distribution exists, we assume  $1/\mu_G < k\mu$  where  $\mu_G$  is the mean of  $G$ .

To analyze this model, we will use an embedded Markov chain approach. Define  $X_n$  as the number in the system as seen by the  $n$ th arrival. Then it is easy to see that  $\{X_n, n \geq 0\}$  is a Markov chain.

To derive the transition probabilities of the Markov chain, it helps to note the relationship

$$X_{n+1} = X_n + 1 - Y_n, \quad n \geq 0$$

where  $Y_n$  denotes the number of departures during the interarrival time between the  $n$ th and  $(n+1)$ st arrival. The transition probabilities can be calculated as

Case (i):  $j > i + 1$ . In this case,  $P_{ij} = 0$ .

Case (ii):  $j \leq i + 1 \leq k$ . In this case,

$P_{ij} = P\{i+1-j \text{ of } i+1 \text{ services are completed in an interarrival time}$

$$= \int_0^\infty P\{i+1-j \text{ of } i+1 \text{ are completed} \mid \text{interarrival time is } t\} dG(t)$$

$$= \int_0^\infty \binom{i+1}{j} (1 - e^{-\mu t})^{i+1-j} (e^{-\mu t})^j dG(t).$$

Case (iii):  $i + 1 \geq j \geq k$ . To evaluate  $P_{ij}$ , in this case, we first note that when all servers are busy the departure process is a Poisson process with rate  $k\mu$ . Hence,

$$\begin{aligned} P_{ij} &= \int_0^{\infty} P\{i+1-j \text{ departures in time } t\} dG(t) \\ &= \int_0^{\infty} e^{-k\mu t} \frac{(k\mu t)^{i+1-j}}{(i+1-j)!} dG(t). \end{aligned}$$

Case (iv):  $i + 1 \geq k > j$ . Conditioning first on the interarrival time and then on the time until there are only  $k$  in the system (call this latter random variable  $T_k$ ) yields

$$\begin{aligned} P_{ij} &= \int_0^{\infty} P\{i+1-j \text{ departures in time } t\} dG(t) \\ &= \int_0^{\infty} \int_0^t P\{i+1-j \text{ departures in } t \mid T_k = s\} \\ &\quad k\mu e^{-k\mu s} \frac{(k\mu s)^{i-k}}{(i-k)!} ds dG(t) \\ &= \int_0^{\infty} \int_0^t \binom{k}{j} (1 - e^{-\mu(t-s)})^{k-j} (e^{-\mu(t-s)})^j \\ &\quad k\mu e^{-k\mu s} \frac{(k\mu s)^{i-k}}{(i-k)!} ds dG(t). \end{aligned}$$

We now can verify by a direct substitution into the equations  $\pi_j = \sum_i \pi_i P_{ij}$  that the limiting probabilities of this Markov chain are of the form

$$\pi_{k-1+j} = c\beta^j, \quad j = 0, 1, \dots.$$

Substitution into any of the equations  $\pi_j = \sum_i \pi_i P_{ij}$  when  $j > k$  yields that  $\beta$  is given as the solution of

$$\beta = \int_0^{\infty} e^{-k\mu t(1-\beta)} dG(t) .$$

The values  $\pi_0, \pi_1, \dots, \pi_{k-2}$ , can be obtained by recursively solving the first  $k - 1$  of the steady-state equations, and  $c$  can then be computed by using  $\sum_0^x \pi_i = 1$ .

If we let  $W_Q^*$  denote the amount of time that a customer spends in queue, then we can show, upon conditioning, that

$$W_Q^* = \begin{cases} 0 & \text{with probability } \sum_0^{k-1} \pi_i = 1 - \frac{c\beta}{1-\beta} \\ \text{Exp}(k\mu(1-\beta)) & \text{with probability } \sum_k^{\infty} \pi_i = \frac{c\beta}{1-\beta} \end{cases}$$

where  $\text{Exp}(k\mu(1-\beta))$  is an exponential random variable with rate  $k\mu(1-\beta)$ .

#### 4. THE FINITE CAPACITY M/G/k

In this section, we consider an M/G/k queuing model having finite capacity  $N$ . That is, a model in which customers, arriving in accordance with a Poisson process having rate  $\lambda$ , enter the system if there are less than  $N$  others present when they arrive, and are then serviced by one of  $k$  servers, each of whom has service distribution  $G$ . Upon entering, a customer will either immediately enter service if at least one server is free or else join the queue if all servers are busy.

Our objective is to obtain an approximation for  $W_Q$ , the average time an entering customer spends waiting in queue. To get started we will make use of the idea that if a (possibly fictitious) cost structure is imposed, so that entering customers are forced to pay money (according to some rule) to the system, then the following identity holds--namely,

time average rate at which the system earns = average arrival rate of entering customers  $\times$  average amount paid by an entering customer.

By choosing appropriate cost rules, many useful formulae can be obtained as special cases. For instance, by supposing

that each customer pays \$1 per unit time while in service, yields  
 average number in service =  $\lambda(1 - P_N)E[S]$ .

Also, if we suppose that each customer in the system pays  $\$x$  per unit time whenever its remaining service time is  $x$ , then we get

$$V = \lambda(1 - P_N)E\left[SW_Q^* + \int_0^S (S - x)dx\right] = \lambda(1 - P_N)[E[S]W_Q + E[S^2]/2]$$

where  $V$  is the (time) average amount of work in the system and where  $W_Q^*$  is a random variable representing the (limiting) amount of time that the  $n$ th entering customer spends waiting in queue.

The above gives us one equation relating  $V$  and  $W_Q^*$  and one approach to obtaining  $W_Q^*$  would be to derive a second equation. An approximate second equation was given by Nozaki-Ross in [3] by means of the following approximation assumption.

#### Approximation Assumption

Given that a customer arrives to find  $i$  busy servers,  $i > 0$ , at the time he enters service the remaining service times of the other customers being served are approximately independent each having the equilibrium service distribution.

Using the above assumption as if it was exactly true, Nozaki-Ross were able to derive a second relationship between  $V$  and  $W_Q^*$  which resulted in an expression for  $W_Q^*$  as a function of  $P_N$ . By approximating  $P_N$  by its known value in the case where the service distribution is exponential, Nozaki-Ross came up with the following approximation for  $W_Q^*$ .

$$W_Q^* = \frac{\frac{E[S]^2}{2E[S]} \sum_{j=k}^{N-1} \frac{(\lambda E[S])^j}{k! k^{j-k}} - (N - k) \frac{E[S](\lambda E[S])^N}{k! k^{N-k}}}{\sum_{j=0}^{k-1} \frac{(\lambda E[S])^j}{j!} + \sum_{j=k}^{N-1} \frac{(\lambda E[S])^j}{k! k^{j-k}} (k - \lambda E[S])}.$$

The idea of an approximation assumption to approximate various quantities of interest of the model M/G/k was also used by Tijms, Van Hoorn, and Federgruen [6]. They used a slightly different approximation assumption to obtain approximations for the steady state probabilities. Other approximations for the M/G/k are also given in Boxma, Cohen, and Huffels [1], and Takahashi [5].

## References

- [1] Boxma, O. J., Cohen, J. W., and Huffels, N.: 1979,  
Operations Research 27, pp. 1115-1127.
- [2] Kelley, F.: 1979, REVERSIBILITY AND STOCHASTIC NETWORKS,  
John Wiley & Sons.
- [3] Nozaki, S. A., and Ross, S. M.: 1978, Journal of Applied  
Probability 15, pp. 826-834.
- [4] Ross, S. M.: 1980, INTRODUCTION TO PROBABILITY MODELS,  
Second Edition, Academic Press.
- [5] Takahashi, Y.: 1977, Journal of the Operations Research  
Society of Japan 20.
- [6] Tijms, H. C., Van Hoorn, M. H., and Federgruen, A.: 1979,  
"Approximations for the Steady State Probabilities in  
the M/G/c queue," Research Report 49, der Vrije  
University, to appear Advances in Applied Probability.

